Scott Berry
Nicholas Erwin
Spring 2021

Reddit Content Analysis Project Proposal

Research Question:

How strongly does a play's change in expected win probability indicate the sentiment of messages directed at referees?

Study Significance:

Referees in sports draw ire from fans on all levels of competition. While referees have messages directed at them throughout the entirety of the games, in particular "big plays" can be heavily criticized and even be cause for concern for referees' safety after the games. Referees from little league to professional level regularly draw death threats, physical violence, and verbal abuse (Jacobs 2020).

This study aims to determine how strongly big plays can affect language directed towards referees and if the sentiment could be at a dangerous level. This knowledge could help predict when a referee might need additional security after a game, as has proved to be necessary many times after competitions worldwide. The findings could support why referee bias leads to questionable decisions if they fear for their safety (Scoppa 2021).

Data Collection:

The data for the study is the comments from the Reddit Game Thread of the NFC Divisional Game between the Tampa Bay Buccaneers and the Green Bay Packers (Reddit 2021). The NFC Divisional game did not have any controversial calls from referees that went viral. This is a significant aspect to our data collection due to the fact that a controversial call from a referee will lead to a higher number of comments about referees during a game. This game thread was selected due to being less than the PRAW maximum of 50,000 comments while still being a recent high-profile game. PRAW is the module to be utilized within a Python script to import all the Reddit comments and their metadata (Boe 2021). Once imported, the comment metadata is reformatted into a JSON structure with comments and their respective bodies, timestamps, and team flairs. The team flairs are included with the goal of associating comments to expected support for home and away teams. The JSON structure is then saved as a file that will be referenced for data analysis.

Data Analysis:

Since the aim is to analyze sentiment towards referees, the comments from the JSON file created from importing the Reddit Game Thread will be filtered down to only the comments that include words associated with referee criticism such as "ref", "penalty", "flag", etc. Depending on

the sample size remaining, this filter may be altered or not applied as general sentiment should mirror referee directed sentiment. If the filter does not reflect referee directed sentiment, the data collection will be skewed and not give an accurate depiction of the general attitude towards referees, which could lead to conflict.

With the remaining comments, VADER sentiment analysis (Hutto & Gilbert 2014) will be performed on the body of each comment using Python scripts. The resulting score will be added alongside the body, timestamp, and team flair keys in the JSON comment structure. Timestamps are used to show when a comment was said about referees during the game and give insight to when big plays may have occurred during a game. Additionally, each comment will be assigned to a play and that play's expected win chance delta (gathered from the ESPN Gamecast page) for the comment author's flair based on its timestamp. The resulting play identifier and change in expected win probability will be added to the comment structure.

With the data collected and properly structured, the sentiment of comments will be plotted against the expected win chance delta. Using these independent and dependent variables, linear regression will be performed to determine the correlation. The results will give the ability to draw conclusions on the expected language and dangers associated with the language that referees experience when calling big plays in football.

References:

Boe, Bryce. "The Python Reddit API Wrapper¶." *PRAW*, 2021, praw.readthedocs.io/en/latest/.

Hutto, CJ, and Eric Gilbert. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

Jacobs, Brittany L., et al. "Exploring Referee Abuse through the Lens of the Collegiate Rugby Coach." *Sport Management Review*, vol. 23, no. 1, 2020, pp. 39–51., doi:10.1016/j.smr.2019.03.004.

Reddit. "r/Nfl - Game Thread: Tampa Bay Buccaneers (11-5) at New Orleans Saints (12-4)." *Reddit*, 2021, www.reddit.com/r/nfl/comments/kzhfxm/game_thread_tampa_bay_buccaneers_115_at_new/.

Scoppa, Vincenzo. "Social Pressure in the Stadiums: Do Agents Change Behavior without Crowd Support?" *Journal of Economic Psychology*, vol. 82, 2021, p. 102344., doi:10.1016/j.joep.2020.102344.